On parameter estimation of hidden telegraph process

RAFAIL Z. KHASMINSKII[1] and YURY A. KUTOYANTS[2]

[1]*Wayne State University, Detroit, USA and Institute of the Information Transmission Problems, Moscow, Russia. E-mail: ab3340@wayne.edu*
[2]*University of Maine, Le Mans, France and National Research University "MPEI", Moscow, Russia. E-mail: kutoyants@univ-lemans.fr*

The problem of parameter estimation is considered for the two-state telegraph process, observed in the white Gaussian observation noise. An online one-step Maximum Likelihood Estimator process is constructed, using a preliminary Method of Moments Estimator. The obtained estimation procedure is shown to be asymptotically normal and asymptotically efficient in the large sample regime.

Keywords: asymptotic efficiency; method of moments estimator; one-step MLE; telegraph process

1. Introduction

In this work, we address the problem of estimating the unknown transition rates λ and μ of a continuous time stationary Markov chain $Y(t), 0 \leq t \leq T$ with two states, y_1 and y_2, and infinitesemal transition matrix

$$\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Trajectory of the chain is observed in a white noise and the unknown two-dimensional parameter $\vartheta = (\lambda, \mu)$ is to be estimated from the observations $X^T = (X(t), 0 \leq t \leq T)$, generated by the equation

$$dX(t) = Y(t)\, dt + dW(t), \qquad \text{subject to } X_0, \tag{1.1}$$

where the initial condition $X(0) = X_0$ is independent of the standard Wiener process $W(t), 0 \leq t \leq T$. We will assume that $\vartheta \in \Theta = (c_0, c_1) \times (c_0, c_1)$ for some known constants $0 < c_0 < c_1$.

Our goal is to construct an *online* estimator process $\vartheta_T^\star = (\vartheta_{t,T}^\star, 0 < t \leq T)$ which, on one hand, is easy to compute, and on the other hand, is asymptotically optimal in an appropriate sense, as $T \to \infty$. The construction is carried out in two steps. First, we introduce a learning interval $[0, T^\delta]$ with $\delta \in (\frac{1}{2}, 1)$ and propose a $T^{\frac{\delta}{2}}$-consistent preliminary estimator using the Method of Moments. Then we improve it up to asymptotic efficiency with the help of a slightly modified one-step MLE procedure.

Observation models, such as the one studied in this paper, are called Hidden Markov Models (HMM) or partially observed systems. HMMs have been extensively studied in the discrete

time setup; see, for example, Elliott *et al.* [5]; Bickel *et al.* [1]; Cappé *et al.* [2] and the references therein. In particular, asymptotic normality of various estimators in the different settings have been established. Asymptotic theory for continuous time models seems to be less explored. Nonlinear filtering and parameter estimation problems for the partially observed continuous time Markov chains are considered, for example, in Elliott *et al.* [5]. Identification problems of partially observed linear processes were studied in, for example, Kutoyants [11]; Dembo and Zeitouni [4]; Zeitouni and Dembo [21]; Kutoyants [12]. The problem of asymptotically efficient estimation for continuous time telegraph process observed on a discrete grid was studied in Iacus and Yoshida [7].

The problem, closest in spirit to ours, is addressed in Chigansky [3], where parameter estimation is considered in the case of hidden finite-state Markov process with continuous time observations. It proves consistency, asymptotic normality and asymptotic efficiency of the MLE under appropriate conditions, which may not be easy to check in general. The case of two-state hidden telegraph process under consideration appears in Chigansky [3] as an example with one-dimensional parameter ($\vartheta = \lambda = \mu$).

The one-step MLE-process proposed here is based on the so-called *Fisher scoring* method. Let us briefly recall its main elements in the case of i.i.d. observations $X^n = (X_1, \ldots, X_n)$, sampled from the density $f(\vartheta, x)$. Suppose that we have a $\sqrt{n}$-consistent estimator $\bar{\vartheta}_n$, that is, $\sqrt{n}(\bar{\vartheta}_n - \vartheta)$ is bounded in probability. Then the estimator $\vartheta_n^\star$

$$\vartheta_n^\star = \bar{\vartheta}_n + \mathbb{I}(\bar{\vartheta}_n)^{-1} n^{-1} \sum_{j=1}^n \frac{\dot{f}(\bar{\vartheta}_n, X_j)}{f(\bar{\vartheta}_n, X_j)}$$

is asymptotically efficient. Here dot stands for the derivative w.r.t. ϑ and $\mathbb{I}(\vartheta)$ is the Fisher information. This estimator was proposed by Fisher [6] and studied by Le Cam [16]. It is also known that the preliminary estimator can have convergence rate slower than $\sqrt{n}$, see Robinson [18] and references therein. Newton–Raphson multi-step estimators were introduced by Kamatani and Uchida [9] in parameter estimation problems for diffusion processes with discrete time observations. Note that the Newton–Raphson procedure is very close to one-step estimation device. In particular, it was shown by Kamatani and Uchida [9] that multi-step Newton–Raphson procedure allows to improve convergence rate of the preliminary estimator up to asymptotic efficiency. The preliminary estimator there is constructed using all available observations.

Let us now illustrate the basic idea behind our MLE-process in this classical discrete time setup. Suppose that $\bar{\vartheta}_N$ is a preliminary estimator constructed by the first $N = [n^\delta]$ observations $X^N = (X_1, \ldots, X_N)$ with $\delta \in (\frac{1}{2}, 1)$. Then the one-step MLE-process $\vartheta_{k,n}^\star$, $N \leq k \leq n$ is given by the following expression

$$\vartheta_{k,n}^\star = \bar{\vartheta}_N + \mathbb{I}_k(\bar{\vartheta}_N)^{-1} k^{-1} \sum_{j=N+1}^k \frac{\dot{f}(\bar{\vartheta}_N, X_j)}{f(\bar{\vartheta}_N, X_j)}, \qquad N \leq k \leq n,$$

where $\mathbb{I}_k(\vartheta)$ is the empirical Fisher information matrix

$$\mathbb{I}_k(\vartheta) = \frac{1}{k} \sum_{j=N+1}^k \frac{\dot{f}(\vartheta, X_j) \dot{f}(\vartheta, X_j)^*}{f(\vartheta, X_j)^2}.$$

Here a^* denotes transpose of a vector (or a matrix) a. For any $\tau \in (0, 1)$ and $k = [n\tau]$ we have

$$\sqrt{k}\left(\vartheta_{k,n}^{\star} - \vartheta\right) \implies \mathcal{N}\left(0, \mathbb{I}(\vartheta)^{-1}\right)$$

as $n \to \infty$. Therefore this estimator-process is asymptotically efficient for all $\tau \in (0, 1]$. Note that its calculation is much simpler than the calculation of the exact MLE.

The MLE-process for the partially observed system (1.1) with observations $X^T = (X_t, 0 \le t \le T)$, proposed in this paper, uses the preliminary estimator $\bar{\vartheta}_{T^\delta}$ based on the observations $X^{T^\delta} = (X_t, 0 \le t \le T^\delta)$ over the initial learning interval $[0, T^\delta]$ with $\delta \in (\frac{1}{2}, 1)$ and is carried out following the similar construction in Kutoyants [15] (see also Kutoyants [14]).

2. Problem statement and auxiliary results

Let us start with description of the MLE for the model under consideration. By the *innovation theorem* (see Liptser and Shiryaev [17], Theorem 7.12), the stochastic process (1.1) admits the representation

$$dX_t = m(t, \vartheta)\, dt + d\bar{W}_t, \quad \text{subject to } X_0, \quad 0 \le t \le T,$$

where $m(t, \vartheta)$ is the conditional expectation

$$m(t, \vartheta) = \mathbf{E}_\vartheta\left[Y(t)|\mathcal{F}_t^X\right] = y_1 \mathbf{P}_\vartheta\left(Y(t) = y_1|\mathcal{F}_t^X\right) + y_2 \mathbf{P}_\vartheta\left(Y(t) = y_2|\mathcal{F}_t^X\right).$$

Here $\mathcal{F}_t^X$ is the σ-algebra of events generated by the observations up to time t, that is, $\mathcal{F}_t^X := \sigma(X_t, 0 \le s \le t)$ and $\bar{W}_t, 0 \le t \le T$ is the *innovation* Wiener process. Let us denote

$$\pi(t, \vartheta) = \mathbf{P}_\vartheta\left(Y(t) = y_1|\mathcal{F}_t^X\right), \qquad \mathbf{P}_\vartheta\left(Y(t) = y_2|\mathcal{F}_t^X\right) = 1 - \pi(t, \vartheta).$$

Then

$$m(t, \vartheta) = y_2 + (y_1 - y_2)\pi(t, \vartheta).$$

The random process $\pi(t, \vartheta), 0 \le t \le T$ satisfies the following equation (see Wonham [20] or Liptser and Shiryaev [17], Theorem 9.1 and equation (9.23) therein)

$$
\begin{aligned}
d\pi(t, \vartheta) = \Big[&\mu - (\lambda + \mu)\pi(t, \vartheta) \\
&+ \pi(t, \vartheta)\left(1 - \pi(t, \vartheta)\right)(y_2 - y_1)\left(y_2 + (y_1 - y_2)\pi(t, \vartheta)\right)\Big]\, dt \qquad (2.1) \\
&+ \pi(t, \vartheta)\left(1 - \pi(t, \vartheta)\right)(y_1 - y_2)\, dX_t.
\end{aligned}
$$

Denote by $\{\mathbf{P}_\vartheta^{(t)}, \vartheta \in \Theta\}$ the family of measures, induced by the random process $X^t = (X_s, 0 \le s \le t)$ from (1.1) on the space of realizations $\mathcal{C}[0, t]$ (continuous functions on $[0, t]$) for different values of the parameter. These measures are equivalent and the likelihood ratio function

$$L\left(\vartheta, X^t\right) = \frac{d\mathbf{P}_\vartheta^{(t)}}{d\mathbf{P}_0^{(t)}}\left(X^t\right), \qquad \vartheta \in \Theta, 0 < t \le T$$

can be written as follows

$$L(\vartheta, X^t) = \exp\left\{ \int_0^t m(s, \vartheta)\, dX_s - \frac{1}{2} \int_0^t m(s, \vartheta)^2\, ds \right\}.$$

Here $\mathbf{P}_0^{(t)}$ is the measure corresponding to X^t with $Y(s) \equiv 0$.

The MLE-process $\hat{\vartheta}_t, 0 < t \le T$ is any root of the equation

$$L(\hat{\vartheta}_t, X^t) = \sup_{\vartheta \in \Theta} L(\vartheta, X^t), \qquad 0 < t \le T. \tag{2.2}$$

Asymptotic behavior of the MLE $\hat{\vartheta}_T$ was studied by Chigansky [3]. It was shown that in the one-dimensional case ($d = 1, \lambda = \mu = \vartheta$) the MLE $\hat{\vartheta}_T$ is consistent, asymptotically normal

$$\sqrt{T}(\hat{\vartheta}_T - \vartheta) \implies \mathcal{N}(0, \mathrm{I}(\vartheta)^{-1})$$

and asymptotically efficient. Here $\mathrm{I}(\vartheta)$ is the Fisher information.

In our model the observations are given by (1.1) and the unknown parameter $\vartheta = (\lambda, \mu)$ has to be estimated from the sample path $X^T = (X_t, 0 \le t \le T)$. We will construct an MLE-process $\vartheta_{t,T}^\star, 0 < t \le T$, such that for any fixed $\tau \in (0, 1]$ and $t = \tau T$ we have

$$\sqrt{t}(\vartheta_{t,T}^\star - \vartheta) \implies \mathcal{N}(0, \mathbb{I}(\vartheta)^{-1}), \qquad T \to \infty.$$

Here $\mathbb{I}(\vartheta)$ is the corresponding Fisher information matrix. Recall that an estimator is asymptotically efficient if it is asymptotically normal with the limit covariance matrix $\mathbb{I}(\vartheta)^{-1}$. The family of measures in the statistical problem under consideration is locally asymptotically normal and therefore it can be shown that the latter implies asymptotic efficiency in the usual minimax sense (see Ibragimov and Has′minskiĭ [8]).

Note that the construction of the MLE-process $\hat{\vartheta}_t, 0 < t \le T$ according to (2.2) and (2.1) is a computationally hard problem, because it requires solving the family of equations (2.1) for all $\vartheta \in \Theta$ and equations (2.2) for all $t \in (0, T]$. Computational complexity can be reduced as follows. First, we construct an estimator $\bar{\vartheta}_T = (\bar{\lambda}_T, \bar{\mu}_T)$ by means of the Method of Moments (see (3.8) below) and show that it is $\sqrt{T}$-consistent, that is,

$$\mathbf{E}_\vartheta \left| \sqrt{T}(\bar{\vartheta}_T - \vartheta) \right|^2 \le C,$$

where the constant $C > 0$ does not depend on T. Then applying this estimator $\bar{\vartheta}_{T^\delta}$ to the observations on the *learning interval* $[0, T^\delta]$ with $\frac{1}{2} < \delta < 1$, we introduce the one-step MLE-process

$$\vartheta_{t,T}^\star = \bar{\vartheta}_{T^\delta} + t^{-1/2} \mathbb{I}_t(\bar{\vartheta}_{T^\delta})^{-1} \Delta_t(\bar{\vartheta}_{T^\delta}, X^t), \qquad T^\delta \le t \le T. \tag{2.3}$$

Here the empirical Fisher information matrix satisfies

$$\mathbb{I}_t(\vartheta) = \frac{1}{t} \int_{T^\delta}^t \dot{m}(\vartheta, s) \dot{m}(\vartheta, s)^*\, ds \longrightarrow \mathbb{I}(\vartheta),$$

as $t \to \infty$, $T^\delta = o(t)$ and

$$\Delta_t(\vartheta, X^t) = \frac{1}{\sqrt{t}} \int_{T^\delta}^t \dot{m}(\vartheta, s)\big[\mathrm{d}X_s - m(\vartheta, s)\,\mathrm{d}s\big]$$

is the vector score-function process.

Computation of this MLE-process uses the solutions $m(\vartheta, t)$ and $\dot{m}(\vartheta, t)$ of the equation (2.1) and of the corresponding equation for $\dot{m}(\vartheta, t)$ only for one value of the parameter, namely for $\vartheta = \bar{\vartheta}_{T^\delta}$. Solving the maximum likelihood equation (2.2) is avoided and the estimator $\vartheta_{t,T}^\star$ can be easily calculated using (2.3). This is the main computational advantage of the proposed estimation procedure. On the other hand, as it was mentioned above, the estimator $\vartheta_{t,T}^\star$ is asymptotically efficient and, in particular, $\vartheta_{T,T}^\star = \vartheta_T^\star$ satisfies

$$\sqrt{T}\big(\vartheta_T^\star - \vartheta\big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \mathbb{I}(\vartheta)^{-1}\big),$$

being therefore asymptotically equivalent to the exact MLE.

Hereafter, we will use the following notations: the derivative w.r.t. parameter ϑ is denoted by a dot, $\dot{m}(\vartheta, t)$ stands for the column vector of the derivatives $\dot{m}_\lambda(\vartheta, t)$, $\dot{m}_\mu(\vartheta, t)$.

Let us recall some relevant well known properties of the (stationary) telegraph process $Y(t), t \geq 0$.

1. The stationary distribution of the process $Y(t)$ is given by

$$\mathbf{P}_\vartheta\big\{Y(t) = y_1\big\} = \frac{\mu}{\lambda + \mu}, \qquad \mathbf{P}_\vartheta\big\{Y(t) = y_2\big\} = \frac{\lambda}{\lambda + \mu}. \tag{2.4}$$

2. Let us denote $P_{ij}(t) = \mathbf{P}_\vartheta\{Y(t) = y_j | Y(0) = y_i\}$, then solving the Kolmogorov equation we obtain

$$P_{11}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad P_{12}(t) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t},$$

$$P_{21}(t) = \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad P_{22}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}. \tag{2.5}$$

It follows from (2.4) and (2.5) that

$$K(s) = \mathbf{E}_\vartheta\big[Y(t)Y(t+s)\big] = \left(\frac{y_1\mu + y_2\lambda}{\lambda + \mu}\right)^2$$

$$+ (y_2 - y_1)^2 \frac{\lambda\mu}{(\lambda + \mu)^2} e^{-(\lambda+\mu)s} = (\bar{Y})^2 + De^{-(\lambda+\mu)s}, \tag{2.6}$$

where

$$\bar{Y} = \mathbf{E}_\vartheta Y(t) = \frac{y_1\mu + y_2\lambda}{\lambda + \mu}, \qquad D = (y_2 - y_1)^2 \frac{\lambda\mu}{(\lambda + \mu)^2}. \tag{2.7}$$

3. Let $\mathcal{F}_t^Y \subset \mathcal{F}$ be the family of σ-algebras of events

$$\{Y(s) = y_i, 0 \leq s \leq t, i = 1, 2\}.$$

It follows from (2.5) that for some constant $K > 0$ and $A < T$ and for all $s > A, t > 0$ the inequality

$$\left|\mathbf{E}_\vartheta\{Y(s+T)Y(t+T)|\mathcal{F}_A^Y\} - \mathbf{E}_\vartheta[Y(s)Y(t)]\right| < Ke^{-(\lambda+\mu)(T-A)} \tag{2.8}$$

holds.

3. Method of moments estimator

Let us first consider the problem of constructing a $\sqrt{T}$-consistent estimator of the unknown parameter ϑ by the Method of Moments. Recall that we observe the stochastic process

$$dX_t = Y(t)\,dt + dW_t, \qquad X_0, \qquad 0 \leq t \leq T, \tag{3.1}$$

in continuous time, where $W_t, 0 \leq t \leq T$ is a standard Wiener process, X_0 is an independent of it random initial condition, and $Y(t), t \geq 0$ is the stationary Markov process with two states y_1 and y_2 and infinitesimal transition matrix

$$\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

The processes $Y(t), t \geq 0$ and $W_t, t \geq 0$ are independent.

For simplicity, we assume that T is an integer and impose the condition

$$\lambda \in [c_0, c_1], \qquad \mu \in [c_0, c_1], \tag{3.2}$$

where c_0 and c_1 are some positive constants. To introduce our preliminary estimator, we need the following objects:

- The function

$$\Phi(x) = \frac{1}{x} - \frac{1}{x^2}(1 - e^{-x}). \tag{3.3}$$

- The statistics

$$\zeta_T = \frac{1}{T}\sum_{i=0}^{T-1}[X_{i+1} - X_i]^2 - 1. \tag{3.4}$$

- The random variable α_T is defined as a solution of the equation

$$\zeta_T = \left(\frac{X_T}{T}\right)^2 + 2\eta_T\Phi(\alpha_T), \tag{3.5}$$

where

$$\eta_T = \left(\frac{X_T}{T} - y_1 \right)\left(y_2 - \frac{X_T}{T} \right). \tag{3.6}$$

- The event $\mathcal{A}_T =$ "*the equation* (3.5) *has a solution* $\alpha_T \in [2c_0, 2c_1]$".
- The random variable

$$\beta_T = \alpha_T \mathbb{1}_{\{\mathcal{A}_T\}} + (c_0 + c_1)\mathbb{1}_{\{\mathcal{A}_T^c\}}. \tag{3.7}$$

Define the estimator $\bar{\vartheta}_T = (\bar{\lambda}_T, \bar{\mu}_T)$ where

$$\bar{\lambda}_T = \frac{\frac{X(T)}{T} - y_1}{y_2 - y_1}\beta_T; \qquad \bar{\mu}_T = \frac{y_2 - \frac{X(T)}{T}}{y_2 - y_1}\beta_T. \tag{3.8}$$

Properties of these estimators are given in the following theorem.

Theorem 1. *Assume the condition* (3.2) *holds*; *then the estimators in* (3.8) *satisfy*

$$\mathbf{E}_\vartheta\left[\sqrt{T}(\bar{\lambda}_T - \lambda) \right]^2 < C, \qquad \mathbf{E}_\vartheta\left[\sqrt{T}(\bar{\mu}_T - \mu) \right]^2 < C \tag{3.9}$$

with a constant $C > 0$, *independent of* T.

The proof is split below into several steps.
The next lemma gives $\sqrt{T}$-consistent estimator for $\bar{Y}$ (see (2.7)).

Lemma 1. *Under the assumption* (3.2), *the estimator* X_T/T *is uniformly consistent for* $\bar{Y}$ *and moreover* :

$$\mathbf{E}_\vartheta\left(\frac{X_T}{T} - \bar{Y} \right)^2 \le \frac{C}{T}, \qquad T > 0 \tag{3.10}$$

with a constant $C > 0$ *independent of* ϑ *and* T.

Proof. Using (2.6), we obtain the following relations

$$\begin{aligned}
\mathbf{E}_\vartheta\left(\frac{X_T}{T} - \bar{Y} \right)^2 &= \mathbf{E}_\vartheta\left| \frac{1}{T}\int_0^T \left[Y(t) - \bar{Y} \right]\mathrm{d}t + \frac{W_T}{T} \right|^2 \\
&= \frac{1}{T} + \frac{1}{T^2}\mathbf{E}_\vartheta\left| \int_0^T \left[Y(t) - \bar{Y} \right]\mathrm{d}t \right|^2 \\
&\le \frac{1}{T}\left(1 + \frac{2\lambda\mu}{(\lambda + \mu)^3}(y_2 - y_1)^2 \right) \le \frac{1}{T}\left(1 + \frac{c_1^2}{4c_0^3}(y_2 - y_1)^2 \right). \qquad \Box
\end{aligned}$$

Corollary. *Existence of consistent estimators for $\frac{\lambda}{\lambda+\mu}$ and $\frac{\mu}{\lambda+\mu}$ follows from (2.4) and Lemma 1. Indeed, from the equality*

$$\bar{Y} = \frac{\lambda}{\lambda+\mu} y_2 + \frac{\mu}{\lambda+\mu} y_1$$

and Lemma 1 we obtain

$$\mathbf{E}_\vartheta \left[\sqrt{T} \left(\frac{T^{-1}X_T - y_1}{y_2 - y_1} - \frac{\lambda}{\lambda+\mu} \right) \right]^2 < C,$$
$$\mathbf{E}_\vartheta \left[\sqrt{T} \left(\frac{y_2 - T^{-1}X_T}{y_2 - y_1} - \frac{\mu}{\lambda+\mu} \right) \right]^2 < C. \tag{3.11}$$

The statistic

$$\frac{X_T}{T} = \frac{1}{T} \int_0^T Y(t)\, dt + \frac{W_T}{T}$$

is a sum of a bounded random variable and an independent normal random variable $N(0, T^{-1})$. Hence, η_T defined in (3.6) satisfies

$$\mathbf{E}_\vartheta \left[\sqrt{T}(\eta_T - D) \right]^2 < C, \tag{3.12}$$

where the constant $C > 0$ does not depend on T and ϑ. The constant D is defined in (2.7).

Note that from the condition (3.2) we have

$$\frac{\lambda\mu}{(\lambda+\mu)^2} > \frac{c_0^2}{4c_1^2}$$

and we easily obtain the bound (3.12) for the estimator

$$\tilde{\eta}_T = \max\left\{ \eta_T, (y_2 - y_1)^2 \frac{c_0^2}{8c_1^2} \right\}. \tag{3.13}$$

Lemma 2. *The following equality holds*

$$\mathbf{E}_\vartheta \zeta_T = \bar{Y}^2 + 2D\Phi(\lambda+\mu) \tag{3.14}$$

and under condition (3.2) we also have

$$\mathbf{E}_\vartheta \left[\sqrt{T}(\zeta_T - \mathbf{E}_\vartheta \zeta_T) \right]^2 < C. \tag{3.15}$$

Proof. By (2.6) and stationarity of the process $Y(t)$, we obtain

$$
\begin{aligned}
\mathbf{E}_\vartheta \zeta_T &= \mathbf{E}_\vartheta [X_1 - X_0]^2 - 1 = \mathbf{E}_\vartheta \int_0^1 \int_0^1 Y(s) Y(t) \, ds \, dt + 1 - 1 \\
&= \bar{Y}^2 + 2D\Phi(\lambda + \mu).
\end{aligned}
\tag{3.16}
$$

Denote

$$
\gamma_i = \int_i^{i+1} Y(t) \, dt; \qquad \Delta W(i) = W_{i+1} - W_i.
$$

Further, the equality

$$
\zeta_T - \mathbf{E}_\vartheta \zeta_T = \frac{1}{T} \sum_{i=0}^{T-1} (\gamma_i^2 - \mathbf{E}_\vartheta \gamma_i^2) + \frac{2}{T} \sum_{i=0}^{T-1} \gamma_i \Delta W(i) + \frac{1}{T} \sum_{i=0}^{T-1} (\Delta W(i)^2 - 1)
$$

implies

$$
\begin{aligned}
\mathbf{E}_\vartheta (\zeta_T - \mathbf{E}_\vartheta \zeta_T)^2 &\le \frac{3}{T^2} \mathbf{E}_\vartheta \left(\sum_{i=0}^{T-1} (\gamma_i^2 - \mathbf{E}_\vartheta \gamma_i^2) \right)^2 + \frac{12}{T^2} \mathbf{E}_\vartheta \left(\sum_{i=0}^{T-1} \gamma_i \Delta W(i) \right)^2 \\
&\quad + \frac{3}{T^2} \mathbf{E}_\vartheta \left(\sum_{i=0}^{T-1} (\Delta W(i)^2 - 1) \right)^2 := 3J_1 + 12J_2 + 3J_3.
\end{aligned}
\tag{3.17}
$$

By stationarity of $Y(t)$

$$
\begin{aligned}
J_1 &= \frac{1}{T^2} \mathbf{E}_\vartheta \left(\sum_{i=0}^{T-1} (\gamma_i^2 - \mathbf{E}_\vartheta \gamma_i^2) \right)^2 = \frac{1}{T^2} \sum_{i=0}^{T-1} \sum_{j=0}^{T-1} \mathbf{E}_\vartheta (\gamma_i^2 - \mathbf{E}_\vartheta \gamma_i^2)(\gamma_j^2 - \mathbf{E}_\vartheta \gamma_j^2) \\
&= \frac{1}{T^2} \sum_{i,j=0}^{T-1} \left\{ \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbf{E}_\vartheta \{ Y(s) Y(t) \right. \\
&\quad \left. \times \mathbf{E}_\vartheta \left[Y(|i-j| + s_1) Y(|i-j| + t_1) | \mathcal{F}_1^Y \right] \} \, ds \, dt \, ds_1 \, dt_1 - (\mathbf{E}_\vartheta \gamma_0^2)^2 \right\}.
\end{aligned}
\tag{3.18}
$$

The bound (2.8) gives

$$
\left| \mathbf{E}_\vartheta \left[Y(|i-j| + s_1) Y(|i-j| + t_1) | \mathcal{F}_1^Y \right] - \mathbf{E}_\vartheta Y(s_1) Y(t_1) \right| \le K e^{-(\lambda + \mu)|j-i|}
$$

and, using (3.18) and (3.2), we obtain

$$
J_1 \le \frac{K}{T^2} \sum_{i,j=0}^{T-1} e^{-(\lambda + \mu)|j-i|} \le \frac{K_1}{T}.
$$

The obvious bounds

$$J_2 = \frac{1}{T^2}\mathbf{E}_\vartheta\left(\sum_{i=0}^{T-1}\gamma_i\Delta W(i)\right)^2 = \frac{1}{T^2}\sum_{i=0}^{T-1}\mathbf{E}_\vartheta\gamma_i^2 = \frac{\mathbf{E}_\vartheta\gamma_0^2}{T} \le \frac{K}{T},$$

$$J_3 = \frac{1}{T^2}\mathbf{E}_\vartheta\left(\sum_{i=0}^{T-1}\left[\Delta W(i)^2 - 1\right]\right)^2 = \frac{1}{T^2}\sum_{i=0}^{T-1}\mathbf{E}_\vartheta\left[\Delta W(i)^2 - 1\right]^2 \le \frac{K}{T}$$

and (3.17) imply the second assertion of the lemma. □

Lemma 3. *The function $\Phi(x)$ (see (3.3)) has the following properties*

$$\lim_{x\to 0+}\Phi(x) = \frac{1}{2}, \tag{3.19}$$

$$\lim_{x\to\infty}\Phi(x) = 0, \tag{3.20}$$

$$\Phi'(x) < 0, \qquad \text{for } x > 0. \tag{3.21}$$

Proof. Expanding (3.3) into power series, we get

$$\Phi(x) = \frac{1}{2} - \frac{x}{3!} + \frac{x^2}{4!} - \frac{x^3}{5!} + \cdots$$

$$\Phi'(x) = -\left(\frac{1}{3!} - \frac{2x}{4!}\right) - \left(\frac{3x^2}{5!} - \frac{4x^3}{6!}\right) - \cdots$$

which give the limits (3.19) and (3.20) and the bound (3.21) for $x < 2$. For $x \ge 2$ it follows from the explicit expression for this derivative

$$\Phi'(x) = \frac{1}{x^2}\left(\frac{2}{x} - 1\right) - \left(\frac{2}{x^3} + \frac{1}{x^2}\right)e^{-x}.$$

Let us consider the equation (3.5) for α_T, where ζ_T and η_T are defined in (3.4) and (3.6) respectively. By Lemma 3 this equation has at most one solution. Recall that on the event $\mathcal{A}_T$, the equation (3.5) does have a solution and consider the statistic β_T defined in (3.7) (here c_0, c_1 are the constants from the condition (3.2)). □

Lemma 4. *Under the condition (3.2), the estimate β_T is $\sqrt{T}$-consistent for $\lambda + \mu$:*

$$\mathbf{E}_\vartheta\left[\sqrt{T}\left(\beta_T - (\lambda + \mu)\right)\right]^2 < C, \tag{3.22}$$

with a constant $C > 0$, independent of ϑ and T.

Proof. It follows from Lemmas 1 and 2 that

$$\zeta_T = \bar{Y}^2 + 2D\Phi(\lambda + \mu) + \varepsilon_1(T). \tag{3.23}$$

Here and below $\varepsilon_i(T), i = 1, 2\ldots$ satisfy

$$\mathbf{E}_\vartheta\left(\sqrt{T}\varepsilon_i(T)\right)^2 < C.$$

By Lemma 1, estimates (3.12), (3.13) and boundedness of $\Phi(x)$ we also obtain

$$\zeta_T = \bar{Y}^2 + 2\tilde{\eta}_T \Phi(\lambda + \mu) + \varepsilon_2(T). \tag{3.24}$$

It follows from (3.5) and (3.24) that on the event $\mathcal{A}_T$

$$2\tilde{\eta}_T \Phi(\alpha_T) = 2\tilde{\eta}_T \Phi(\lambda + \mu) + \varepsilon_3(T).$$

Since $\tilde{\eta}_T$ is bounded away from zero by a positive constant (see Corollary to Lemma 1), the latter equality and Lemma 3 imply that on $\mathcal{A}_T$

$$\Phi(\alpha_T) - \Phi(\lambda + \mu) = \varepsilon_4(T).$$

Therefore by Lemma 3, we obtain

$$\mathbf{E}_\vartheta\left\{\mathbb{1}_{\{\mathcal{A}_T\}}\sqrt{T}\left(\beta_T - (\lambda + \mu)\right)\right\}^2 < C. \tag{3.25}$$

If $\omega \in \mathcal{A}_T^c$ then the equation

$$\bar{Y}^2 + 2D\Phi(\lambda + \mu) + \varepsilon_3(T) = \left(\frac{X(T)}{T}\right)^2 + 2\tilde{\eta}_T \Phi(x) \tag{3.26}$$

has no solution $x \in [2c_0, 2c_1]$.

It follows from (3.26), Lemma 1 and the corollary that the equation

$$\Phi(x) = \Phi(\lambda + \mu) + \varepsilon_4(T)$$

has no solution for $x \in [2c_0, 2c_1]$. This means that

$$\left[\bar{Y}^2 + 2D\Phi(\lambda + \mu) + \varepsilon_3 - (\bar{Y} + \varepsilon_0)^2\right](2D + 2\varepsilon_5)^{-1} \notin \left[\Phi(2c_0), \Phi(2c_1)\right],$$

where we used the notations $\varepsilon_0 = T^{-1}X_T - \bar{Y}$ and $\varepsilon_5 = \tilde{\eta}_T - D$.

Hence, we can write

$$\mathcal{A}_T^c \subset \left\{\left|\varepsilon_0(T)\right| > \alpha_1\right\} \cup \left\{\left|\varepsilon_3(T)\right| > \alpha_2\right\} \cup \left\{\left|\varepsilon_5(T)\right| > \alpha_3\right\}$$

for some positive constants α_1, α_2 and α_3, which do not depend on T. This inclusion implies

$$\mathbf{P}\left(\mathcal{A}_T^c\right) \leq \mathbf{P}\left\{\left|\varepsilon_0(T)\right| > \alpha_1\right\} + \mathbf{P}\left\{\left|\varepsilon_3(T)\right| > \alpha_2\right\} + \mathbf{P}\left\{\left|\varepsilon_5(T)\right| > \alpha_3\right\} < \frac{C}{T}$$

and the claim of Lemma 4 holds by (3.25). $\square$

Proof of Theorem 1. Let us now apply obtained results to prove $\sqrt{T}$-consistency of the estimators defined in (3.9). To this end, the obvious equality

$$\bar{\lambda}_T - \lambda = \beta_T \frac{\frac{X(T)}{T} - y_1}{y_2 - y_1} - \beta_T \frac{\lambda}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu}(\beta_T - (\lambda + \mu)),$$

the bounds (3.11) and Lemma 4 give

$$\mathbf{E}_\vartheta \left(\sqrt{T}(\bar{\lambda}_T - \lambda) \right)^2 \leq 2\mathbf{E}_\vartheta \left[\sqrt{T}\beta_T \left(\frac{\frac{X(T)}{T} - y_1}{y_2 - y_1} - \frac{\lambda}{\lambda + \mu} \right) \right]^2$$

$$+ 2\left(\frac{\lambda}{\lambda + \mu} \right)^2 \mathbf{E}_\vartheta \left[\sqrt{T}(\beta_T - (\lambda + \mu)) \right]^2 < C.$$

The second inequality in (3.9) is shown similarly and thus the estimator $\bar{\vartheta}_T = (\bar{\lambda}_T, \bar{\mu}_T)$ is $\sqrt{T}$-consistent. $\qquad\square$

4. One-step MLE

Our next goal is to construct the asymptotically efficient MLE-process for the parameter $\vartheta = (\lambda, \mu) \in \Theta$. We do it in two steps. First, we define the preliminary estimator $\bar{\vartheta}_{T^\delta} = (\bar{\lambda}_{T^\delta}, \bar{\mu}_{T^\delta})$, obtained by applying the estimator from the previous section to the observations $X^{T^\delta} = (X_t, 0 \leq t \leq T^\delta)$ on the learning interval $[0, T^\delta]$, with $\delta \in (\frac{1}{2}, 1)$. By Theorem 1, this estimator satisfies the condition:

$$\sup_{\vartheta \in K} T^\delta \mathbf{E}_\vartheta |\bar{\vartheta}_{T^\delta} - \vartheta|^2 \leq C,$$

with a constant $C > 0$ independent of T and $\vartheta \in \Theta$. Let us introduce the additional condition (recall that $\Theta = (c_0, c_1) \times (c_0, c_1)$):

$\mathcal{M}(N)$. *There exists $N \geq 2$, such that*

$$\frac{c_0}{(y_1 - y_2)^2} > \frac{2N + 9}{8}. \tag{4.1}$$

With the preliminary estimator $\bar{\vartheta}_{T^\delta}$ at hand, we propose one-step MLE:

$$\vartheta^\star_{t,T} = \bar{\vartheta}_{T^\delta} + t^{-1}\mathbb{I}_t(\bar{\vartheta}_{T^\delta})^{-1} \int_{T^\delta}^t \dot{m}(\bar{\vartheta}_{T^\delta}, s)\left[dX_s - m(\bar{\vartheta}_{T^\delta}, s) \, ds \right], \tag{4.2}$$

based on the slightly modified score-function

$$\Delta_t(\vartheta, X^t) = \frac{1}{\sqrt{t}} \int_0^t \dot{m}(\vartheta, s)\left[dX_s - m(\vartheta, s) \, ds \right], \qquad T^\delta \leq t \leq T.$$

Here we defined the vector

$$\dot{m}(\vartheta, s) = (y_1 - y_2)\frac{\partial \pi(s, \vartheta)}{\partial \vartheta} = (y_1 - y_2)\left(\frac{\partial \pi(t, \vartheta)}{\partial \lambda}, \frac{\partial \pi(t, \vartheta)}{\partial \mu}\right)^*$$

and the empirical Fisher information matrix $\mathbb{I}_t(\vartheta)$

$$\mathbb{I}_t(\vartheta) = \frac{1}{t}\int_{T^\delta}^t \dot{m}(\vartheta, s)\dot{m}(\vartheta, s)^* \, \mathrm{d}s \longrightarrow \mathbb{I}(\vartheta), \qquad t \to \infty,$$

where the limit holds by the Law of Large Numbers; $\mathbb{I}(\vartheta)$ is the Fisher information matrix

$$\mathbb{I}(\vartheta) = (y_1 - y_2)^2 \mathbf{E}_\vartheta \frac{\partial \pi(s, \vartheta)}{\partial \vartheta}\frac{\partial \pi(s, \vartheta)^*}{\partial \vartheta}.$$

The stochastic process $\frac{\partial \pi(s,\vartheta)}{\partial \vartheta}$ has relevant ergodic properties (see the proof of Lemma 6 below) and we assume here that $\frac{\partial \pi(s,\vartheta)}{\partial \vartheta}$ has invariant distribution. Therefore, the expectation above does not depend on s.

Let us change the variable $\tau = tT^{-1} \in [0, 1]$ and introduce the stochastic process $\vartheta_T^\star(\tau)$, $\tau_\delta \le \tau \le 1$, where $\vartheta_T^\star(\tau) = \vartheta_{\tau T, T}^\star$ and $\tau_\delta = T^{\delta-1} \to 0$. Below ϑ_0 denotes the true value of the parameter.

Theorem 2. *Suppose $\vartheta_0 \in \Theta$, $\delta \in (\frac{1}{2}, 1)$ and the condition $\mathcal{M}(2)$ holds. Then the one-step MLE-process is consistent: for any $v > 0$ and any $\tau \in (0, 1]$*

$$\mathbf{P}_{\vartheta_0}\{|\vartheta_T^\star(\tau) - \vartheta_0| > v\} \to 0 \tag{4.3}$$

as $T \to \infty$. Moreover, it is asymptotically normal

$$\sqrt{\tau T}(\vartheta_T^\star(\tau) - \vartheta_0) \implies \mathcal{N}(0, \mathbb{I}(\vartheta_0)^{-1}). \tag{4.4}$$

Proof. Let us denote the partial derivatives

$$\dot{\pi}_\lambda(t, \vartheta) = \frac{\partial \pi(t, \vartheta)}{\partial \lambda}, \qquad \dot{\pi}_\mu(t, \vartheta) = \frac{\partial \pi(t, \vartheta)}{\partial \mu}, \qquad \ddot{\pi}_{\lambda,\lambda}(t, \vartheta) = \frac{\partial^2 \pi(t, \vartheta)}{\partial \lambda^2},$$

and so on. $\square$

Lemma 5. *Suppose $\vartheta_0 \in \Theta$ and $N > 1$. If the condition*

$$\frac{c_0}{(y_1 - y_2)^2} > \frac{N + 3}{4} \tag{4.5}$$

holds, then

$$\sup_{\vartheta \in \Theta} \mathbf{E}_{\vartheta_0}\left(|\dot{\pi}_\lambda(t, \vartheta)|^N + |\dot{\pi}_\mu(t, \vartheta)|^N\right) < C_1, \tag{4.6}$$

and, if the condition

$$\frac{c_0}{(y_1 - y_2)^2} > \frac{2N + 9}{8} \tag{4.7}$$

holds, then

$$\sup_{\vartheta \in \Theta} \mathbf{E}_{\vartheta_0}\left(\left|\ddot{\pi}_{\lambda,\lambda}(t, \vartheta)\right|^N + \left|\ddot{\pi}_{\lambda,\mu}(t, \vartheta)\right|^N + \left|\ddot{\pi}_{\mu,\mu}(t, \vartheta)\right|^N\right) < C_2. \tag{4.8}$$

Here the constants $C_1 > 0, C_2 > 0$ do not depend on t.

Proof. For simplicity, we write

$$\dot{\pi}_\lambda(t, \vartheta) = \dot{\pi}_\lambda, \qquad \dot{\pi}_\mu(t, \vartheta) = \dot{\pi}_\mu, \qquad \pi(t, \vartheta) = \pi.$$

Taking the derivative of

$$d\pi = \left[\mu - (\lambda + \mu)\pi - \pi(1 - \pi)(y_1 - y_2)\left(y_2 + (y_1 - y_2)\pi\right)\right]dt \\ + \pi(1 - \pi)(y_1 - y_2)\,dX_t, \tag{4.9}$$

we obtain the equations

$$d\dot{\pi}_\lambda = -\pi\,dt - \dot{\pi}_\lambda\left[\lambda + \mu + (1 - 2\pi)(y_1 - y_2)\left[y_2 + (y_1 - y_2)\pi\right]\right. \\ \left. + \pi(1 - \pi)(y_1 - y_2)^2\right]dt + \dot{\pi}_\lambda(1 - 2\pi)(y_1 - y_2)\,dX(t), \tag{4.10}$$

$$d\dot{\pi}_\mu = [1 - \pi]\,dt - \dot{\pi}_\mu\left[\lambda + \mu + (1 - 2\pi)(y_1 - y_2)\left[y_2 + (y_1 - y_2)\pi\right]\right. \\ \left. + \pi(1 - \pi)(y_1 - y_2)^2\right]dt + \dot{\pi}_\mu(1 - 2\pi)(y_1 - y_2)\,dX(t). \tag{4.11}$$

Using the standard arguments it can be shown that the stochastic process $\pi(t, \vartheta)$ has continuous derivatives w.r.t. λ and μ with probability 1. If we denote the true value of the parameters by ϑ_0 and $\pi(t, \vartheta_0) = \pi^o$ etc., then these equations with $\vartheta = \vartheta_0$ read

$$d\dot{\pi}_\lambda^o = -\pi^o\,dt - \dot{\pi}_\lambda^o\left[\lambda_0 + \mu_0 + \pi^o(1 - \pi^o)(y_1 - y_2)^2\right]dt \\ + \dot{\pi}_\lambda^o(1 - 2\pi^o)(y_1 - y_2)\,d\bar{W}(t), \tag{4.12}$$

$$d\dot{\pi}_\mu^o = [1 - \pi^o]\,dt - \dot{\pi}_\mu^o\left[\lambda_0 + \mu_0 + \pi^o(1 - \pi^o)(y_1 - y_2)^2\right]dt \\ + \dot{\pi}_\mu^o(1 - 2\pi)(y_1 - y_2)\,d\bar{W}(t), \tag{4.13}$$

where

$$d\pi^o = \left[\mu_0 - (\lambda_0 + \mu_0)\pi^o\right]dt + \pi^o(1 - \pi^o)(y_1 - y_2)\,d\bar{W}(t). \tag{4.14}$$

This system is linear with respect to $\dot{\pi}_\lambda$ and $\dot{\pi}_\mu$ and it can be written as follows ($x_t = \pi^o$, $y_t = \dot{\pi}_\lambda^o$, $z_t = \dot{\pi}_\mu^o$, $a = \lambda_0 + \mu_0$, $b = y_1 - y_2$)

$$dx_t = [\mu_0 - ax_t]\,dt + bx_t(1 - x_t)\,d\bar{W}_t, \tag{4.15}$$

$$dy_t = -x_t\,dt - \left[a + b^2 x_t(1 - x_t)\right]y_t\,dt + b(1 - 2x_t)y_t\,d\bar{W}_t, \tag{4.16}$$

$$dz_t = [1 - x_t]\,dt - \left[a + b^2 x_t(1 - x_t)\right]z_t\,dt + b(1 - 2x_t)z_t\,d\bar{W}_t. \tag{4.17}$$

Note that as $\lambda_0 > 0$ and $\mu_0 > 0$, the process $\pi(t, \vartheta_0) = x_t \in (0, 1)$ is ergodic with two reflecting boundaries at 0 and 1. Therefore the process $\pi^o(t, \vartheta_0)$ is ergodic with the invariant density

$$f(\vartheta_0, x) = \frac{[x(1-x)]^{\frac{2(\mu_0 - \lambda_0)}{(y_1 - y_2)^2} - 2}}{G(\vartheta_0)} \cdot \exp\left\{-\frac{2\mu_0 + 2(\lambda_0 - \mu_0)x}{(y_1 - y_2)^2 x(1 - x)}\right\}$$

$$= \frac{[x(1-x)]^{\gamma(\mu_0 - \lambda_0) - 2}}{G(\vartheta_0)} \exp\left\{-\frac{\gamma\mu_0}{x} - \frac{\gamma\lambda_0}{1 - x}\right\},$$

where we denote $\gamma = 2(y_1 - y_2)^{-2}$ and define the normalizing constant

$$G(\vartheta_0) = \int_0^1 [x(1-x)]^{\gamma(\mu_0 - \lambda_0) - 2} \exp\left\{-\frac{\gamma\mu_0}{x} - \frac{\gamma\lambda_0}{1 - x}\right\}dx.$$

The processes y_t and z_t admit the following expressions

$$y_t = -\int_0^t \exp\left\{-\int_v^t \left[a + b^2 x_s(1 - x_s) - \frac{b^2}{2}(1 - 2x_s)^2\right]ds \right.$$

$$\left. + b\int_v^t (1 - 2x_s)\,d\bar{W}_s\right\} x_v\,dv, \tag{4.18}$$

$$z_t = \int_0^t \exp\left\{-\int_v^t \left[a + b^2 x_s(1 - x_s) - \frac{b^2}{2}(1 - 2x_s)^2\right]ds \right.$$

$$\left. + b\int_v^t (1 - 2x_s)\,d\bar{W}_s\right\}[1 - x_v]\,dv. \tag{4.19}$$

Let us put $\mathrm{x}_s = \frac{1}{2} - x_s$. Then we have

$$x_s(1 - x_s) - \frac{1}{2}(1 - 2x_s)^2 = -3\mathrm{x}_s^2 + \frac{1}{4}$$

and

$$y_t = \int_0^t \left(\mathrm{x}_v - \frac{1}{2}\right)e^{-(a + \frac{b^2}{4})(t - v)}\exp\left\{3b^2 \int_v^t \mathrm{x}_s^2\,ds + 2b\int_v^t \mathrm{x}_s\,d\bar{W}_s\right\}dv.$$

To estimate the moments $\mathbf{E}_{\vartheta_0}|y_t|^N$, we note that $|x_v - \frac{1}{2}| \le \frac{1}{2}$ and use the Hölder inequality

$$\left(\int_0^t \left|f(v)g(v)\right| dv\right)^N \le \left(\int_0^t \left|f(v)\right|^{\frac{N}{N-1}} dv\right)^{N-1} \int_0^t \left|g(v)\right|^N dv$$

with $f(v) = \exp\{-a(t - v)\varepsilon\}$ and

$$g(v) = \exp\left\{-\left(a(1 - \varepsilon) + \frac{b^2}{4}\right)(t - v) + 3b^2 \int_v^t x_s^2 \, ds + 2b \int_v^t x_s \, d\bar{W}_s\right\},$$

where $\varepsilon > 0$. This yields the estimate

$$\mathbf{E}_{\vartheta_0}|y_t|^N \le C(N, \varepsilon) \int_0^t e^{-N(a(1-\varepsilon)+\frac{b^2}{4})(t-v)} \mathbf{E}_{\vartheta_0} e^{3Nb^2 \int_v^t x_s^2 \, ds + 2Nb \int_v^t x_s \, d\bar{W}_s} \, dv,$$

where the constant $C(N, \varepsilon) > 0$ does not depend on t. Further, we can write

$$\mathbf{E}_{\vartheta_0} \exp\left\{3Nb^2 \int_v^t x_s^2 \, ds + 2Nb \int_v^t x_s \, d\bar{W}_s\right\}$$

$$= \mathbf{E}_{\vartheta_0}\left(\exp\left\{2Nb \int_v^t x_s \, d\bar{W}_s - 2N^2 b^2 \int_v^t x_s^2 \, ds\right\}\right.$$

$$\left.\times \exp\left\{Nb^2(2N + 3) \int_v^t x_s^2 \, ds\right\}\right)$$

$$\le \exp\left\{\frac{Nb^2}{4}(2N + 3)(t - v)\right\}$$

because $x_s^2 \le 1/4$, and

$$\mathbf{E}_{\vartheta_0} \exp\left\{2Nb \int_v^t x_s \, d\bar{W}_s - 2N^2 b^2 \int_v^t x_s^2 \, ds\right\} = 1.$$

Therefore,

$$\mathbf{E}_{\vartheta_0}|y_t|^N \le C(N, \varepsilon) \int_0^t e^{-N(a(1-\varepsilon)+\frac{b^2}{4}-\frac{b^2}{4}(2N+3))(t-v)} \, dv$$

$$= C(N, \varepsilon) \int_0^t e^{-N(a(1-\varepsilon)-\frac{b^2}{2}(N+1))(t-v)} \, dv.$$

We see that if

$$\frac{\lambda_0 + \mu_0}{(y_1 - y_2)^2} > \frac{1}{2} + \frac{N}{2},$$

then $\mathbf{E}_{\vartheta_0}|y_t|^N \le C$. In particular, if we set $N = 2$ in the condition (4.5) and choose a sufficiently small $\varepsilon > 0$, we get the bound

$$\sup_{\vartheta_0 \in \Theta} \mathbf{E}_{\vartheta_0} \left| \frac{\partial \pi(t, \vartheta_0)}{\partial \lambda} \right|^2 \le C, \tag{4.20}$$

where the constant $C > 0$ does not depend on t.

We also need to estimate the derivatives (4.10), (4.11) for the values $\vartheta \ne \vartheta_0$. The equation for $\dot{\pi}_\lambda$ becomes

$$\begin{aligned}
\mathrm{d}\dot{\pi}_\lambda = &-\pi \, \mathrm{d}t - \dot{\pi}_\lambda \big[\lambda + \mu + (1 - 2\pi)(y_1 - y_2)^2 \big(\pi - \pi^0 \big) \\
&+ \pi(1 - \pi)(y_1 - y_2)^2 \big] \, \mathrm{d}t + \dot{\pi}_\lambda (1 - 2\pi)(y_1 - y_2) \, \mathrm{d}\bar{W}(t).
\end{aligned} \tag{4.21}$$

Hence if we put $a = \lambda + \mu$, $y_t = \dot{\pi}_\lambda$ and $b = y_1 - y_2$, we obtain the equation

$$\begin{aligned}
\mathrm{d}y_t = &-x_t \, \mathrm{d}t - \big[a + b^2 (1 - 2x_t)\big(x_t - x_t^0\big) + b^2 x_t (1 - x_t) \big] y_t \, \mathrm{d}t \\
&+ b(1 - 2x_t) y_t \, \mathrm{d}\bar{W}_t.
\end{aligned}$$

The solution of this equation can be written explicitly, similarly to (4.18), but with additional term $b^2 (1 - 2x_t)(x_t - x_t^0)$ in the exponent. This term satisfies the inequality

$$(1 - 2x_t)\big(x_t - x_t^0\big) \ge -1.$$

Hence by calculations as above, $\mathbf{E}_{\vartheta_0}|y_t|^2$ will be bounded, if

$$\frac{\lambda + \mu}{(y_1 - y_2)^2} > \frac{3}{2} + \frac{N}{2}.$$

For the second derivative $\ddot{\pi} = \ddot{\pi}_{\lambda,\lambda}(t, \vartheta)$ we obtain similar bounds for the moments as follows. The equation for $\ddot{\pi}$ is

$$\begin{aligned}
\mathrm{d}\ddot{\pi} = &-y_t \big[2 - 2b^2 y_t \big(x_t - x_t^0\big) + 2b^2 y_t (1 - 2x_t) \big] \, \mathrm{d}t - 2b y_t^2 \, \mathrm{d}\bar{W}_t \\
&- \ddot{\pi} \big[a + b^2 (1 - 2x_t)\big(x_t - x_t^0\big) + b^2 x_t (1 - x_t) \big] \, \mathrm{d}t + b\ddot{\pi}(1 - 2x_t) \, \mathrm{d}\bar{W}_t.
\end{aligned}$$

Let us write it as

$$\mathrm{d}\ddot{\pi} = A(t) \, \mathrm{d}t + B(t) \, \mathrm{d}\bar{W}_t - \ddot{\pi}\big[a + C(t)\big] \, \mathrm{d}t + \ddot{\pi}_t D(t) \, \mathrm{d}\bar{W}_t$$

using the obvious notations. Its solution satisfies

$$\frac{\partial^2 \pi(t, \vartheta)}{\partial \lambda^2} = \int_0^t e^{-\int_v^t [a + C(s) - \frac{1}{2} D(s)^2] \, \mathrm{d}s + \int_v^t D(s) \, \mathrm{d}\bar{W}_s} \big[A(v) \, \mathrm{d}v + B(v) \, \mathrm{d}\bar{W}_v \big].$$

The bound

$$C(s) - \frac{D(s)^2}{2} = b^2(1 - 2x_s)\left(x_s - x_s^0\right) + \frac{b^2}{2}\left[2x_s(1 - x_s) - (1 - 2x_s)^2\right]$$

$$\geq -b^2 - 3b^2\left(x - \frac{1}{2}\right)^2 + \frac{b^2}{4} = -\frac{3b^2}{4} - 3b^2\left(x - \frac{1}{2}\right)^2 \geq -\frac{3b^2}{2}$$

holds since $(x - \frac{1}{2})^2 \leq \frac{1}{4}$. Therefore,

$$a - \frac{3}{2}b^2 - \frac{2N + 3}{4}b^2 = a - \frac{9}{4}b^2 - \frac{N}{2}b^2$$

and if

$$\frac{\lambda + \mu}{(y_1 - y_2)^2} > \frac{9}{4} + \frac{N}{2},$$

then we obtain

$$\mathbf{E}_{\vartheta_0}\left|\frac{\partial^2 \pi(t, \vartheta)}{\partial \lambda^2}\right|^N < C.$$

Hence under the condition (4.7), we have

$$\sup_{\vartheta \in \Theta} \mathbf{E}_{\vartheta_0}\left|\frac{\partial^2 \pi(t, \vartheta)}{\partial \lambda^2}\right|^N < C.$$

Similar estimates can be obtained for other derivatives and the claimed assertion follows. $\square$

Lemma 6. *The solutions* (x_t, y_t, z_t) *of the equations* (4.15)–(4.17) *have the following ergodic properties*

$$\frac{1}{T}\int_0^T \dot{m}_\lambda(t, \vartheta_0)^2 \, dt = \frac{b^2}{T}\int_0^T y_t^2 \, dt \longrightarrow I_{11}(\vartheta_0),$$

$$\frac{1}{T}\int_0^T \dot{m}_\lambda(t, \vartheta_0)\dot{m}_\mu(t, \vartheta_0) \, dt = \frac{b^2}{T}\int_0^T y_t z_t \, dt \longrightarrow I_{12}(\vartheta_0),$$

$$\frac{1}{T}\int_0^T \dot{m}_\mu(t, \vartheta_0)^2 \, dt = \frac{b^2}{T}\int_0^T z_t^2 \, dt \longrightarrow I_{22}(\vartheta_0),$$

where convergence is in the mean square.

Proof. The proof of existence of the invariant measure can be found in Chigansky [3], Section 4.2. Note that the equations (4.15)–(4.17) do not coincide with those in Chigansky [3], where the case $\lambda = \mu$ and $y_1 = 1$, $y_2 = 0$ is considered; however the same arguments apply to the system of equations (4.15)–(4.17).

The strong mixing coefficient $\alpha(t)$ for ergodic diffusion process (4.15) satisfies the estimate

$$\alpha(t) < e^{-c|t|},$$

as shown by Veretennikov [19]. The conditions, under which this bound holds, can be checked by applying the following transformation to the equation (4.15):

$$\xi_t = g(x_t), \qquad g(x) = \int_{1/2}^{x} \frac{dv}{bv(1-v)}, \qquad x \in (0,1).$$

The obtained process solves the stochastic differential equation

$$d\xi_t = A(\xi_t)\,dt + dW_t, \qquad \xi_0 = g(x_0), \qquad 0 \le t \le T,$$

whose drift coefficient $A(\cdot)$ satisfies the assumptions needed in Veretennikov [19].

Now to verify the convergence

$$\mathbf{E}_{\vartheta_0}\left(\frac{1}{T}\int_0^T y_t^2\,dt - \frac{1}{T}\int_0^T \mathbf{E}_{\vartheta_0} y_t^2\,dt\right)^2 = \mathbf{E}_{\vartheta_0}\left(\frac{1}{T}\int_0^T \left[y_t^2 - \mathbf{E}_{\vartheta_0} y_t^2\right]dt\right)^2 \longrightarrow 0 \quad (4.22)$$

we can apply the result of the following lemma. $\square$

Lemma 7. *Let $\{Y_t, t > 0\}$ be a stochastic process with zero mean, satisfying*

$$\mathbf{E}|Y_t|^{m(2k-1)} < C_1, \qquad \int_0^{\infty} t^{k-1}\left[\alpha(t)\right]^{(m-2)/m}\,dt < C_2,$$

for some $m > 2$ and $k \ge 1$, where $\alpha(t)$ is the strong mixing coefficient. Then

$$\mathbf{E}\left|\int_0^T Y_t\,dt\right|^{2k} \le C_3 T^k.$$

Proof. For the proof see Lemma 2.1 in Khasminskii [10]. $\square$

The limit (4.22) is obtained by setting $Y_t = y_t^2 - \mathbf{E}_{\vartheta_0} y_t^2$, $m = 3$ and $k = 1$.

Let us now verify consistency of the one-step MLE-process. To this end, we have

$$\mathbf{P}_{\vartheta_0}\left\{\left|\vartheta_T^{\star}(\tau) - \vartheta_0\right| > \nu\right\}$$

$$\le \mathbf{P}_{\vartheta_0}\left\{|\bar{\vartheta}_{T^\delta} - \vartheta_0| > \frac{\nu}{2}\right\}$$

$$+ \mathbf{P}_{\vartheta_0}\left\{\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\left[dX_s - m(\bar{\vartheta}_{T^\delta}, s)\,ds\right]\right| > \frac{\nu}{2}\right\}.$$

By the Theorem 1, the first term on the right satisfies

$$\mathbf{P}_{\vartheta_0}\left\{|\bar{\vartheta}_{T^\delta} - \vartheta_0| > \frac{\nu}{2}\right\} \le \frac{4}{\nu^2}\mathbf{E}_{\vartheta_0}|\bar{\vartheta}_{T^\delta} - \vartheta_0|^2 \le \frac{C}{\nu^2 T^\delta} \to 0.$$

The second term can be estimated as follows

$$\mathbf{P}_{\vartheta_0}\left\{\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)[dX_s-m(\bar{\vartheta}_{T^\delta},s)\,ds]\right|>\frac{\nu}{2}\right\}$$

$$\leq\mathbf{P}_{\vartheta_0}\left\{\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,d\bar{W}_s\right|>\frac{\nu}{4}\right\}$$

$$+\mathbf{P}_{\vartheta_0}\left\{\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\Delta m(\bar{\vartheta}_{T^\delta},s)\,ds\right|>\frac{\nu}{4}\right\},$$

where $\Delta m(\bar{\vartheta}_{T^\delta},s)=m(\vartheta_0,s)-m(\bar{\vartheta}_{T^\delta},s)$. We can write

$$\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,d\bar{W}_s\right|\leq\frac{\|\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}\|}{T^\gamma}\left|\frac{1}{T^{\delta-\gamma}}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,d\bar{W}_s\right|,$$

where $0<\gamma<\delta-\frac{1}{2}$. Hence

$$\mathbf{P}_{\vartheta_0}\left\{\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,d\bar{W}_s\right|>\frac{\nu}{4}\right\}$$

$$\leq\mathbf{P}_{\vartheta_0}\left\{\frac{1}{T^{\delta-\gamma}}\left|\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,d\bar{W}_s\right|>\frac{\sqrt{\nu}}{2}\right\}$$

$$+\mathbf{P}_{\vartheta_0}\left\{\frac{\|\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}\|}{T^\gamma}>\frac{\sqrt{\nu}}{2}\right\}\longrightarrow 0,$$

as $T\to\infty$, because

$$\mathbf{P}_{\vartheta_0}\left\{\frac{1}{T^{\delta-\gamma}}\left|\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,d\bar{W}_s\right|>\frac{\sqrt{\nu}}{2}\right\}$$

$$\leq\frac{1}{\nu T^{2\delta-2\gamma}}\mathbf{E}_{\vartheta_0}\int_{T^\delta}^{T}\left|\dot{m}(\bar{\vartheta}_{T^\delta},s)\right|^2\,ds\leq\frac{C}{\nu T^{2\delta-2\gamma-1}}\to 0.$$

Recall that $2\delta-2\gamma-1>0$; then

$$\mathbf{P}_{\vartheta_0}\left\{\left|\frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\Delta m(\bar{\vartheta}_{T^\delta},s)\,ds\right|>\frac{\nu}{4}\right\}$$

$$\leq\mathbf{P}_{\vartheta_0}\left\{\frac{1}{\tau T^{1-\gamma}}\left|\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\int_0^1\dot{m}(\vartheta_v,s)^*\,dv\,ds(\bar{\vartheta}_{T^\delta}-\vartheta_0)\right|>\frac{\sqrt{\nu}}{2}\right\}$$

$$+\mathbf{P}_{\vartheta_0}\left\{\frac{\|\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}\|}{T^\gamma}>\frac{\sqrt{\nu}}{2}\right\}\longrightarrow 0$$

as $T\to\infty$, because $\bar{\vartheta}_{T^\delta}-\vartheta_0=O(T^{-\delta/2})$ and other terms are bounded in probability (here $\vartheta_v=\vartheta_0+v(\bar{\vartheta}_{T^\delta}-\vartheta_0)$).

To prove (4.4), let us write

$$\sqrt{\tau T}\big(\vartheta_T^\star(\tau) - \vartheta_0\big)$$

$$= \sqrt{\tau T}(\hat{\vartheta}_{T^\delta} - \vartheta_0) + \frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\sqrt{\tau T}} \int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\, d\bar{W}_s$$

$$+ \frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\sqrt{\tau T}} \int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\big[m(\vartheta_0, s) - m(\bar{\vartheta}_{T^\delta}, s)\big]\, ds,$$

and note that

$$\mathbf{E}_{\vartheta_0} \left| \frac{1}{\sqrt{\tau T}} \int_{T^\delta}^{\tau T} \big[\dot{m}(\bar{\vartheta}_{T^\delta}, s) - \dot{m}(\vartheta_0, s)\big]\, d\bar{W}_s \right|^2$$

$$\leq \frac{1}{\tau T} \int_{T^\delta}^{\tau T} \mathbf{E}_{\vartheta_0} \big|\dot{m}(\bar{\vartheta}_{T^\delta}, s) - \dot{m}(\vartheta_0, s)\big|^2\, ds \longrightarrow 0$$

as $T \to \infty$. By the central limit theorem (see, e.g., Theorem 1.19 in Kutoyants [13]) the convergence in distribution

$$\frac{1}{\sqrt{\tau T}} \int_{T^\delta}^{\tau T} \dot{m}(\vartheta_0, s)\, d\bar{W}_s \quad \Longrightarrow \quad \mathcal{N}\big(0, \mathbb{I}(\vartheta_0)\big)$$

holds, since by the Law of Large Numbers

$$\frac{1}{\tau T} \int_{T^\delta}^{\tau T} \dot{m}(\vartheta_0, s)\dot{m}(\vartheta_0, s)^*\, dt \longrightarrow \mathbb{I}(\vartheta_0).$$

Further, let us denote $\hat{v}_{T^\delta} = \sqrt{\tau T}(\hat{\vartheta}_{T^\delta} - \vartheta_0)$, then we can write

$$\hat{v}_{T^\delta} + \frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\sqrt{\tau T}} \int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\big[m(\vartheta_0, s) - m(\bar{\vartheta}_{T^\delta}, s)\big]\, ds$$

$$= \mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1} \left(\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta}) - \frac{1}{\tau T} \int_0^1 \int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\dot{m}(\vartheta_r, s)^*\, dr\, ds \right) \hat{v}_{T^\delta},$$

where $\vartheta_r = \bar{\vartheta}_{T^\delta} + r(\bar{\vartheta}_{T^\delta} - \vartheta_0)$. The expression

$$\dot{m}(\vartheta_r, s) = \dot{m}(\bar{\vartheta}_{T^\delta}, s) + \int_0^1 \ddot{m}(\vartheta_q, s)\, dq\, (\bar{\vartheta}_{T^\delta} - \vartheta_0)$$

and the equality

$$\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta}) = \frac{1}{\tau T} \int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\dot{m}(\bar{\vartheta}_{T^\delta}, s)^*\, ds$$

give

$$\hat{v}_{T^\delta} + \frac{\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}}{\sqrt{\tau T}} \int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\big[m(\vartheta_0, s) - m(\bar{\vartheta}_{T^\delta}, s)\big] \mathrm{d}s$$

$$= \sqrt{\tau T}|\bar{\vartheta}_{T^\delta} - \vartheta_0|^2 O(1) = T^{\frac{1}{2}-\delta} O(1) \longrightarrow 0,$$

as $T \to \infty$.

Let us verify that the Fisher information matrix is nondegenerate. To this end, it suffices to show invertibility of the matrix

$$\mathbb{J}(\vartheta_0) = \begin{pmatrix} \mathbf{E}_{\vartheta_0} \tilde{y}_t^2 & \mathbf{E}_{\vartheta_0} \tilde{y}_t \tilde{z}_t \\ \mathbf{E}_{\vartheta_0} \tilde{y}_t \tilde{z}_t & \mathbf{E}_{\vartheta_0} \tilde{z}_t^2 \end{pmatrix},$$

where $\tilde{y}_t, \tilde{z}_t$ are stationary solutions of (4.16) and (4.17) respectively. If this matrix is degenerate, then

$$\mathbf{E}_{\vartheta_0} \tilde{y}_t^2 \mathbf{E}_{\vartheta} \tilde{z}_t^2 = (\mathbf{E}_{\vartheta_0} \tilde{y}_t \tilde{z}_t)^2. \tag{4.23}$$

By the Cauchy–Schwarz inequality

$$(\mathbf{E}_{\vartheta_0} \tilde{y}_t \tilde{z}_t)^2 \le \mathbf{E}_{\vartheta_0} \tilde{y}_t^2 \mathbf{E}_{\vartheta_0} \tilde{z}_t^2$$

with equality if and only if $\tilde{z}_t = c\tilde{y}_t$ with some constant $c \ne 0$. Therefore in the case of equality, we have $\mathbf{E}_{\vartheta_0}(c\tilde{y}_t - \tilde{z}_t)^2 = 0$.

Introduce a new process $\tilde{v}_t = c\tilde{y}_t - \tilde{z}_t$ as a solution of the equation

$$\mathrm{d}\tilde{v}_t = \big[\tilde{x}_t(1-c) - 1\big]\mathrm{d}t - \big[a + b^2\tilde{x}_t(1-\tilde{x}_t)\big]\tilde{v}_t\,\mathrm{d}t + b(1-2\tilde{x}_t)\tilde{v}_t\,\mathrm{d}\tilde{W}_t,$$

where $\tilde{v}_t$ and $\tilde{x}_t$ are stationary.

Further, following Section 4 of Chigansky [3], we can write

$$\tilde{v}_t = \tilde{v}_0 e^{-at} + \int_0^t e^{-a(t-s)}\big[\tilde{x}_s(1-c) - 1\big]\mathrm{d}s - b^2 \int_0^t e^{-a(t-s)}\tilde{x}_s(1-\tilde{x}_s)\tilde{v}_s\,\mathrm{d}s$$

$$+ b\int_0^t e^{-a(t-s)}(1-2\tilde{x}_s)\tilde{v}_s\,\mathrm{d}\tilde{W}_s.$$

Hence,

$$\mathbf{E}_{\vartheta_0}\left(\int_0^t e^{-a(t-s)}\big[\tilde{x}_s(1-c) - 1\big]\mathrm{d}s\right)^2$$

$$\le 4\big(1 + e^{-2at}\big)\mathbf{E}_{\vartheta_0}\tilde{v}_t^2$$

$$+ \frac{4b^4}{a}\int_0^t e^{-a(t-s)}\frac{1}{16}\mathbf{E}_{\vartheta_0}\tilde{v}_s^2\,\mathrm{d}s + 4b^2\int_0^t e^{-2a(t-s)}\mathbf{E}_{\vartheta_0}\tilde{v}_s^2\,\mathrm{d}s \le C\mathbf{E}_{\vartheta_0}\tilde{v}_t^2$$

with some constant $C > 0$ which does not depend on t. By stationarity $\mathbf{E}_{\vartheta_0}\tilde{v}_t^2$ does not depend on t and hence non-degeneracy of the matrix $\mathbb{J}(\vartheta_0)$ follows, if we can show that

$$\lim_{t\to\infty} \mathbf{E}_{\vartheta_0}\left(\int_0^t e^{-a(t-s)}\big[\tilde{x}_s(1-c)-1\big]\,\mathrm{d}s\right)^2 > 0,$$

for all c. The stochastic process

$$\zeta_t = \int_0^t e^{-a(t-s)}\big[\tilde{x}_s(1-c)-1\big]\,\mathrm{d}s$$

is the solution of the equation

$$\frac{\mathrm{d}\zeta_t}{\mathrm{d}t} = -a\zeta_t + \tilde{x}_t(1-c)-1, \qquad \zeta_0 = 0.$$

Elementary calculations show that for all ϑ_0 and c

$$\lim_{t\to\infty} \mathbf{E}_{\vartheta_0}\zeta_t^2 = \frac{\mathbf{E}_{\vartheta_0}[\tilde{\pi}_0(1-c)-1]^2}{a^2} > 0,$$

where $\tilde{\pi}_0$ is a random variable with the density of the invariant distribution of the stochastic process $\pi(t,\vartheta_0)$. Thus the Fisher information matrix $\mathbb{I}(\vartheta_0)$ is nondegenerate for all $\vartheta_0 \in \Theta$.

Note that, by the Theorem 2, the limit covariance matrix of the one-step MLE-process coincides with the covariance of the asymptotically efficient MLE (see Chigansky [3]), therefore $\vartheta_T^\star(\tau)$ is asymptotically efficient as well.

4.1. Discussion

In this section, we dwell on several possible generalizations of the obtained results.

Using the same arguments as in Kutoyants [15] it is possible to prove the uniform consistency: for any $\nu > 0$

$$\lim_{T\to\infty} \mathbf{P}_{\vartheta_0}\left\{ \sup_{\tau_\delta \leq \tau \leq 1} \big|\vartheta_T^\star(\tau) - \vartheta_0\big| \geq \nu \right\} = 0.$$

Moreover it can be shown that the normalized one-step MLE-process

$$\eta_T(\tau) = \tau\sqrt{T}\mathbb{I}(\vartheta_0)^{-1/2}\big(\vartheta_T^\star(\tau) - \vartheta_0\big), \qquad \tau_* \leq \tau \leq 1,$$

converges to the two-dimensional standard Wiener process $W(\tau)$, $\tau_* \leq \tau \leq 1$, where τ_* is any value satisfying $\tau_* \in (0,1]$. For the details, see the proof of a similar result in Kutoyants [15].

The condition $\mathcal{M}(2)$ can probably be relaxed, but it needs a special study. The finiteness of the Fisher information follows from Chigansky [3] and the condition $\mathcal{M}(2)$ is basically used to control the matrix of the second derivatives.

4.2. Markov chain with $d > 2$ states

In the case of Markov chain $Y(t), t \geq 0$ with $d > 2$ states, the only essential difficulty in our approach is the construction of the preliminary estimator. For two states construction of a suitable preliminary estimator, is relatively easy and it is not immediately clear how to extend it to the more general setup. If we have some $\sqrt{T}$ consistent estimator $\bar{\vartheta}_T$ of the unknown parameter $\vartheta \in \Theta \subset R^m$, then the corresponding one-step MLE-process is

$$\vartheta_t^\star = \bar{\vartheta}_{T^\delta} + t^{-1}\mathbb{I}_t(\bar{\vartheta}_{T^\delta})^{-1} \int_{T^\delta}^t \dot{m}(\bar{\vartheta}_{T^\delta}, s)\big[\mathrm{d}X_s - m(\bar{\vartheta}_{T^\delta}, s)\,\mathrm{d}s\big], \qquad T^\delta \leq t \leq T,$$

where $\delta \in (\frac{1}{2}, 1)$. Asymptotic efficiency

$$\sqrt{\tau T}\big(\vartheta_{\tau T}^\star - \vartheta_0\big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \mathbb{I}(\vartheta_0)^{-1}\big),$$

can be shown, applying the arguments from this paper. Note that the estimator $\vartheta_T^\star$ is computationally simpler than the MLE $\hat{\vartheta}_T$.

For the discrete time hidden Markov models with $d \geq 2$ states, the MLE studied in for example, Bickel *et al.* [1], can be used as the preliminary estimator for the one-step MLE-process, constructed as in our paper. The obtained *online* estimator of the unknown parameter will have asymptotically optimal properties.

4.3. Two-step MLE-process

The learning interval $[0, T^\delta]$ with $\delta \in (\frac{1}{2}, 1)$ is negligible with respect to the observations time T. It can be made even shorter, if we use two-step MLE-process approach, as it was proposed by Kutoyants [15]. This modification uses the learning interval $[0, T^\delta)$ with $\delta \in (\frac{1}{3}, \frac{1}{2}]$. The procedure is as follows. First, we obtain the preliminary estimator $\bar{\vartheta}_{T^\delta}$ by applying our Method of Moments estimator to the observations $X^{T^\delta} = (X_s, 0 \leq s \leq T^\delta)$. By (3.9)

$$T^\delta \mathbf{E}_{\vartheta_0} |\bar{\vartheta}_{T^\delta} - \vartheta_0|^2 \leq C,$$

where the constant $C > 0$ does not depend on T.

Then we introduce the second preliminary estimator-process

$$\bar{\vartheta}_{t,T} = \bar{\vartheta}_{T^\delta} + t^{-1/2}\mathbb{I}_t(\bar{\vartheta}_{T^\delta})^{-1}\Delta_t\big(\bar{\vartheta}_{T^\delta}, X^t\big), \qquad t \in \big[T^\delta, T\big].$$

The two-step MLE-process is

$$\vartheta_{t,T}^{\star\star} = \bar{\vartheta}_{t,T} + t^{-1/2}\mathbb{I}_t(\bar{\vartheta}_{T^\delta})^{-1}\Delta_t\big(\bar{\vartheta}_{T^\delta}, \bar{\vartheta}_{t,T}, X^t\big), \qquad t \in \big[T^\delta, T\big],$$

where

$$\Delta_t\big(\vartheta_1, \vartheta_2, X^t\big) = \frac{1}{\sqrt{t}} \int_{T^\delta}^t \dot{m}(\vartheta_1, s)\big[\mathrm{d}X_s - m(\vartheta_2, s)\,\mathrm{d}s\big], \qquad t \in \big[T^\delta, T\big].$$

It can be shown that for all $\tau \in (0, 1]$ and $t = \tau T$ we have the asymptotic normality of the estimator $\vartheta_T^{\star\star}(\tau) = \vartheta_{\tau T, T}^{\star\star}$:

$$\sqrt{\tau T}\big(\vartheta_T^{\star\star}(\tau) - \vartheta_0\big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \mathbb{I}(\vartheta_0)^{-1}\big). \tag{4.24}$$

The first step of the proof is to study the expression

$$T^{\frac{\gamma}{2}}(\bar{\vartheta}_{\tau T, T} - \vartheta_0) = T^{\frac{\gamma}{2}}(\bar{\vartheta}_{T^\delta} - \vartheta_0) + T^{-1+\frac{\gamma}{2}}\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}\Delta_{\tau T, T}(\vartheta_0, X^t)$$

$$+ T^{-1+\frac{\gamma}{2}}\mathbb{I}_{\tau T}(\bar{\vartheta}_{T^\delta})^{-1}\int_{T^\delta}^{\tau T} \dot{m}(\bar{\vartheta}_{T^\delta}, s)\big[m(\vartheta_0, s) - m(\bar{\vartheta}_{T^\delta}, s)\big]\,\mathrm{d}s.$$

We can show that if $\frac{2}{3} < \gamma < 2\delta$, then $T^{\frac{\gamma}{2}}(\bar{\vartheta}_{\tau T, T} - \vartheta_0)$ is bounded in probability and we have (4.24).

The proof is omitted, being similar in spirit to that in Kutoyants [15]. Note however that such two-step MLE-process lacks the advantage of being computationally efficient: its construction requires generating the family of filters $m(\bar{\vartheta}_{s,T}, s)$ for all $\bar{\vartheta}_{s,T}, s \in [T^\delta, t]$, while for one-step MLE-process, $m(\vartheta, s)$ is to be computed only for the single value $\vartheta = \bar{\vartheta}_{T^\delta}$.

4.4. Kalman filter

The proposed device is universal and can be applied to construct asymptotically efficient estimators (estimator-processes) in other models. Let us consider the following simple example of the partially observed system

$$\mathrm{d}X_t = (\lambda + \mu Y_t)\,\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 = 0,$$

$$\mathrm{d}Y(t) = -aY(t)\,\mathrm{d}t + \mathrm{d}V_t, \qquad Y_0,$$

where $W_t, t \geq 0$ and $V_t, t \geq 0$ are two independent Wiener processes and $a > 0$ is a known constant. The unknown parameter $\vartheta = (\lambda, \mu) \in \Theta = (c_0, c_1) \times (c_3, c_4)$ has to be estimated, using the sample $X^T = (X_t, 0 \leq t \leq T)$. Let us define $n(\vartheta, t) = \mathbf{E}_\vartheta(Y(t)|\mathcal{F}_t^X)$ and $m(\vartheta, t) = \lambda + \mu n(\vartheta, t)$, where $\mathcal{F}_t^X = \sigma\{X_s, 0 \leq s \leq t\}$. The conditional expectation $n(\vartheta, t)$ satisfies the Kalman–Bucy filtering equations

$$\mathrm{d}n(\vartheta, t) = -an(\vartheta, t)\,\mathrm{d}t + \mu\gamma_t(\mu)\big[\mathrm{d}X_t - \big(\lambda + \mu n(\vartheta, t)\big)\,\mathrm{d}t\big], \qquad n(\vartheta, 0),$$

$$\frac{\partial\gamma_t(\mu)}{\partial t} = -2a\gamma_t(\mu) - \gamma_t(\mu)^2\mu^2 + 1, \qquad \gamma_0, \qquad 0 \leq t \leq T$$

and the log-likelihood function is given by

$$L\big(\vartheta, X^T\big) = \int_0^T m(\vartheta, t)\,\mathrm{d}X_t - \frac{1}{2}\int_0^T m(\vartheta, t)^2\,\mathrm{d}t, \qquad \vartheta \in \Theta.$$

Calculation of the MLE $\hat{\vartheta}_T$ requires $m(\vartheta, t)$ for all $\vartheta \in \Theta, t \in [0, T]$ (see, e.g., Section 3.1 in Kutoyants [13]) and hence we are faced with the computational problem of the same complexity

as for the model considered in this paper. One-step procedure can reduce the computational load, while attaining the optimal performance in the large sample limit.

A reasonable choice of the preliminary estimator is $\bar{\vartheta}_T = (\bar{\lambda}_T, \bar{\mu}_T)$ with

$$\bar{\lambda}_T = \frac{X_T}{T}, \qquad \bar{\mu}_T = \sqrt{\frac{\zeta_T}{K}},$$

where we denoted

$$\zeta_T = \max(\bar{\zeta}_T, 0), \qquad \bar{\zeta}_T = \frac{1}{T} \sum_{i=0}^{T-1} [X_{i+1} - X_i - \bar{\lambda}_T]^2 - 1, \qquad K = \mathbf{E}\left(\int_0^1 Y_s^* \, \mathrm{d}s\right)^2.$$

Here Y_s^* is the stationary process satisfying the same equation as $Y(s)$. It can be shown that

$$\mathbf{E}_{\vartheta_0} \left| \sqrt{T}(\bar{\vartheta}_T - \vartheta_0) \right|^2 < C.$$

Introduce the learning interval $[0, T^\delta]$ with $\delta \in (\frac{1}{2}, 1)$ and the one-step MLE

$$\vartheta_T^\star = \bar{\vartheta}_{T^\delta} + \mathbb{I}_T(\bar{\vartheta}_{T^\delta})^{-1} T^{-1} \int_{T^\delta}^T \dot{m}(\bar{\vartheta}_{T^\delta}, t) \big[\mathrm{d}X_t - \big(\bar{\lambda}_{T^\delta} + \bar{\mu}_{T^\delta} n(\bar{\vartheta}_{T^\delta}, t)\big) \, \mathrm{d}t\big].$$

Using the same arguments as above, we can verify the convergence

$$\sqrt{T}\big(\vartheta_T^\star - \vartheta_0\big) \implies \mathcal{N}\big(0, \mathbb{I}(\vartheta_0)^{-1}\big).$$

Acknowledgment

References

[1] Bickel, P.J., Ritov, Y. and Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26** 1614–1635. MR1647705

[2] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. New York: Springer. MR2159833

[3] Chigansky, P. (2009). Maximum likelihood estimator for hidden Markov models in continuous time. *Stat. Inference Stoch. Process.* **12** 139–163. MR2511676

[4] Dembo, A. and Zeitouni, O. (1986). Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Process. Appl.* **23** 91–113. MR0866289

[5] Elliott, R.J., Aggoun, L. and Moor, J.B. (1995). *Hidden Markov Models*. New York: Springer.

[6] Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Camb. Philos. Soc.* **22** 700–725.

[7]   Iacus, S. and Yoshida, N. (2009). Estimation for discretly observed telegraph process. *Theory Probab. Math. Statist.* **78** 37–47.

[8]   Ibragimov, I.A. and Has′minskiĭ, R.Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. New York-Berlin: Springer. MR0620321

[9]   Kamatani, K. and Uchida, M. (2015). Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat. Inference Stoch. Process.* **18** 177–204. MR3348584

[10] Khasminskii, R.Z. (1966). On stochastic processes defined by differential equations with small parameter. *Theory Probab. Appl.* **11** 240–259.

[11] Kutoyants, Y.A. (1984). *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann.

[12] Kutoyants, Y.A. (1994). *Identification of Dynamical Systems with Small Noise*. *Mathematics and Its Applications* **300**. Dordrecht: Kluwer Academic. MR1332492

[13] Kutoyants, Y.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. London: Springer. MR2144185

[14] Kutoyants, Y.A. (2014). Approximation of the solution of the backward stochastic differential equation. Small noise, large sample and high frequency cases. *Proc. Steklov Inst. Math.* **287** 133–154. MR3484327

[15] Kutoyants, Y.A. (2015). On multi-step MLE-processes for ergodic diffusion. *Stochastic Process. Appl.* To appear.

[16] Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 129–156. Berkeley and Los Angeles: Univ. California Press. MR0084918

[17] Liptser, R.S. and Shiryaev, A.N. (2001). *Statistics of Random Processes*. *I*, 2nd ed. New York: Springer.

[18] Robinson, P.M. (1988). The stochastic difference between econometric statistics. *Econometrica* **56** 531–548. MR0946120

[19] Veretennikov, A.Y. (1987). Bounds for the mixing rate in the theory of stochastic equations. *Theory Probab. Appl.* **32** 273–281.

[20] Wonham, W.M. (1965). Some applications of stochastic differential equations to optimal nonlinear filtering. *J. Soc. Indust. Appl. Math. Ser. A Control* **2** 347–369. MR0186472

[21] Zeitouni, O. and Dembo, A. (1988). Exact filters for the estimation of the number of transitions of finite-state continuous-time Markov processes. *IEEE Trans. Inform. Theory* **34** 890–893. MR0966763